Hadoop Assignmnet 2

Section 5: Using relational data stores with hadoop

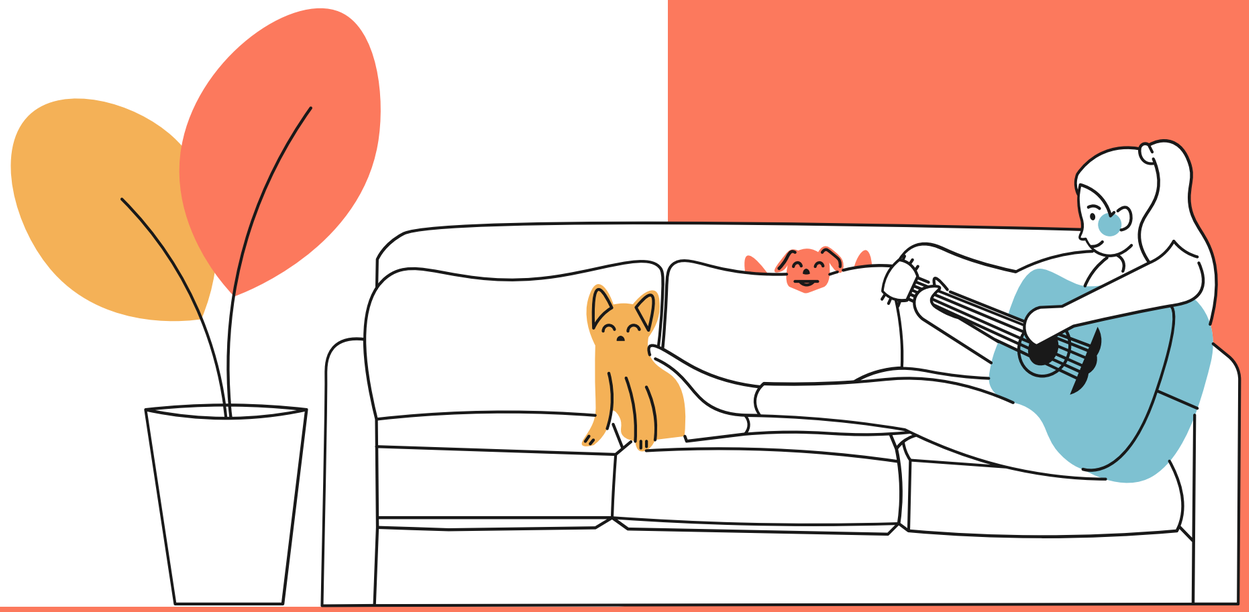
My notes:

Hadoop Section 5.txt - Notepad

File Edit Format View Help

```
What is Hive?

mapreduce or tez

OLAP queries

No transactions, no updates, inserts or deletes

like mysql IT TAKES TIME!

Upload via CSV.

create view topmovieId's AS
select movieId, count(movieID) as ratingCount
From ratings
Group By movieID
Order By ratingCount Desc;

Go google Hive Commands

"IF NOT EXISTS" "DROP VIEW"

How does Hive Work?

Schema On Read

Ways to use Hive

What is MySQL

Sqoop (NB)

Incremental imports

Use sqoop to import data from MySQL to HFDS/Hive

Use Sqoop to export data from Hadoop to MySQL
```

Ln 37, Col 46 | 100% | Windows (CRLF) | UTF-8



Apache Hive

What is Hive?

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data and makes querying and analyzing easy.

OLAP QUERIES

- **Rolling up** to country level:

```
SELECT COUNT(visits), SUM(sales)
GROUP BY country
```

Country	visits	sales
USA	4	$50
Canada	1	0

- **“Slice”** by browser

```
SELECT COUNT(visits), SUM(sales)
GROUP BY country
HAVING browser = “FF”
```

Country	visits	sales
USA	2	$10
Canada	0	0

- **Top** browsers by sales

```
SELECT SUM(sales), COUNT(visits)
GROUP BY browser
ORDER BY sales
```

Browser	sales	visits
Chrome	$25	2
Safari	$15	1
FF	$10	2

Online analytical processing (OLAP) is a system for performing multi-dimensional analysis at high speeds on large volumes of data.

- No Transactions
- No Updates
- No Deletes

HIVE COMMANDS

DDL Command	Function
CREATE	It is used to create a table or Database
SHOW	It is used to show Database, Table, Properties, etc
ALTER	It is used to make changes to the existing table
DESCRIBE	It describes the table columns
TRUNCATE	Used to permanently truncate and delete the rows of table
DELETE	Deletes the table data, but, can be restored

Practical

Sqoop is a command-line interface application for transferring data between relational databases and Hadoop.

Use Sqoop to import data from MySQL to HFDS/Hive
Use Sqoop to export data from Hadoop to MySQL

```
mysql> exit
Bye

C:\>
```

Thanks for listening

Enrico Dreyer 31210783